

Genomic signatures associated with small-scale farming landscapes in wild bee populations: potential adaptive response under high gene flow

Octavio Paulo

octavio.paulo@fc.ul.pt

Universidade de Lisboa <https://orcid.org/0000-0001-5408-5212>

André Henriques

Universidade de Lisboa <https://orcid.org/0000-0003-2800-046X>

Paulo Sousa

Martin-Luther University Halle-Wittenberg <https://orcid.org/0000-0002-2333-6493>

Sara Silva

Universidade de Lisboa

Davide Galhofa

Universidade de Lisboa <https://orcid.org/0000-0002-3514-4296>

Luísa Carvalheiro

Universidade Federal de Goiás

Thomas Wood

Naturalis, Biodiversity Center

Orianne Rollin

ITSAP - Institut de l'Abeille <https://orcid.org/0000-0002-1455-9302>

Article

Keywords: Landscape genomics; RAD-seq; wild bees; genotype-environment association; candidate loci; pollinator conservation

Posted Date: October 19th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10753096/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: There is no duality of interest

1 **Abstract**

2 Intensified agricultural landscapes can influence both genetic diversity and signatures
3 of selection in wild bees, yet the genomic consequences of retaining small-scale
4 farming (SSF) landscapes remain poorly understood. We used RAD sequencing to
5 investigate population structure, genetic diversity, signatures of selection, and
6 genotype-environment associations in three widespread wild bee species sampled
7 across contrasting SSF and low-agricultural (LoA) landscapes in Portugal: *Andrena*
8 *flavipes* Panzer, 1799, *Bombus terrestris* (Linnaeus, 1758), and *Lasioglossum*
9 *malachurum* (Kirby, 1802).

10 Across species, genetic diversity was high and comparable between SSF and LoA
11 landscapes, genetic differentiation was low, population structure was weak, and
12 isolation by distance was not detected, suggesting high connectivity among sampling
13 locations. Complementary selection detection approaches identified conservative
14 sets of candidate SNPs associated with land-use contrasts and composite gradients of
15 landscape configuration. Functional annotation of overlapping candidate SNPs
16 identified genes involved in signaling, cellular communication, ion transport,
17 metabolism, transcriptional regulation, and sensory or stress-related processes. Our
18 results suggest that SSF landscapes may maintain genetic diversity while also being
19 associated with genomic signatures that provide a basis for future studies of wild bee
20 adaptation.

21 **Keywords**

22 Landscape genomics; RAD-seq; wild bees; genotype-environment association;
23 candidate loci; pollinator conservation

24 **1 | Introduction**

25 Climate and land-use change are major drivers of biodiversity loss (Finn *et al.*, 2023).
26 Wild bees have declined in abundance, diversity, and distribution (Soroye *et al.*, 2020;
27 Janousek *et al.*, 2023), largely due to agricultural expansion, habitat simplification,
28 pesticide exposure, climate change, and parasites and pathogens (McNeil *et al.*, 2020;
29 Raven and Wagner, 2021; Outhwaite *et al.*, 2022). Understanding how agricultural
30 landscapes affect wild bees genetic diversity and adaptive potential is therefore a
31 central conservation challenge (Marshall *et al.*, 2024).

32 Agricultural intensification can reduce floral and nesting resources, alter flowering
33 availability, and impose novel selective pressures on wild bee (Grab *et al.*, 2019; Raven
34 and Wagner, 2021; Outhwaite *et al.*, 2022). By simplifying landscapes and increasing
35 habitat fragmentation, intensified agricultural land use can also reduce population
36 connectivity, restrict gene flow, and generate detectable genetic structure across
37 multiple taxa, although the magnitude of these effects depends on species' dispersal
38 ability and ecological traits (Jha, 2015; Haugen *et al.*, 2024; Ho *et al.*, 2025). However,
39 some wild bee species can persist in human-modified landscapes while showing
40 phenotypic and allele-frequency shifts associated with local environmental and land-
41 use conditions, indicating that adaptive divergence may occur despite high gene flow

42 and weak genetic differentiation (Theodorou *et al.*, 2018, 2021; Jackson *et al.*, 2018,
43 2020; Kelemen and Rehan, 2021; Glück *et al.*, 2022).

44 Small-scale farming (SSF) landscapes (farms less than 5 ha in size) remain
45 understudied in consideration of wild bee landscape genomics. This is a relevant gap
46 since SSF landscapes represent a substantial component (62.8%) of European
47 agricultural systems (Eurostat, 2022). In contrast to large-scale intensive agriculture,
48 SSF landscapes are characterized by mosaics of small fields, diverse crops, semi-
49 natural elements, field margins, and dispersed rural settlements (Hass *et al.*, 2018;
50 Gutiérrez-Briceño *et al.*, 2023; Albrecht *et al.*, 2023). This heterogeneity can promote
51 pollinator abundance and diversity by enhancing connectivity, nesting opportunities,
52 and floral resource availability (Brandt *et al.*, 2017; Sirami *et al.*, 2019). At the same
53 time, SSF landscapes remain subject to anthropogenic pressures, including fertilizer
54 and pesticide application, soil disturbance, and mechanized land management. Thus,
55 SSF landscapes may simultaneously maintain connectivity and generate spatially
56 heterogeneous selection.

57 From an evolutionary perspective, this dual role may produce high gene flow and weak
58 population structure while allowing selection to drive localized allele-frequency shifts
59 at a limited number of genomic regions ((Barghi *et al.*, 2020; Yeaman, 2022). Detecting
60 these signals is challenging because gene flow homogenizes most of the genome.
61 Combining complementary genome-scan approaches and prioritizing candidates
62 supported across multiple methods can therefore provide a conservative strategy for
63 identifying loci and genomic regions consistent with putative selection (Lotterhos and
64 Whitlock, 2014; François *et al.*, 2016; Ahrens *et al.*, 2018).

65 Here, we investigate genomic signatures associated with SSF landscapes in three wild
66 bee species: *Andrena flavipes* Panzer, 1799, *Bombus terrestris* (Linnaeus, 1758), and
67 *Lasioglossum malachurum* (Kirby, 1802). These species are widespread habitat and
68 dietary generalists, but differ in body size, social and nesting organization, and
69 dispersal capacity, providing a comparative framework to test for shared or species-
70 specific genomic responses to land-use heterogeneity. Using RAD sequencing, we
71 tested four hypotheses: (i) SSF landscapes will maintain genetic diversity and exhibit
72 weak population structure relative to landscapes with low proportion of agriculture
73 (LoA); (ii) despite high gene flow, a restricted number of candidate loci will show
74 repeated support across complementary outlier approaches, indicating localized
75 genomic responses to selection; (iii) genotype-environment associations and
76 functional annotations will differ among species, reflecting species-specific responses
77 to landscape configuration.

78 **2 | Materials and Methods**

79 **2.1 | Study area, target species and sampling design**

80 The study was conducted in a 2,220 km² coastal small-scale farming landscape in
81 western Portugal. The region is characterized by strong Atlantic influence, mild climatic
82 conditions, and a heterogeneous land-use mosaic composed of a mixed of small
83 agricultural fields, fruit production, vineyards, orchards, rural settlements, semi-
84 natural vegetation, pine forests, and nearby mountain systems.

85 These conditions support high biodiversity and agricultural productivity (Viana *et al.*,
86 2021), making the region suitable for investigating how heterogeneous SSF landscapes

87 shape wild bee genetic diversity, connectivity, and genomic variation. In addition,
88 recent climate change projections identify this area as important for climatic refugia
89 and biodiversity persistence under future scenarios, reinforcing the conservation
90 relevance of understanding evolutionary processes in this landscape (Araújo, 2022).

91 Three widespread wild bee species were selected: *Andrena flavipes*, *Bombus terrestris*,
92 and *Lasioglossum malachurum*. These species are common in Portugal but differ in
93 body size, sociality, nesting ecology, and dispersal capacity (Baldock *et al.*, 2018). *A.*
94 *flavipes* is a medium-sized, ground-nesting solitary bee (Sousa *et al.*, 2023; Szczepko-
95 Morawiec *et al.*, 2024); *B. terrestris* is a large, social bumblebee with high dispersal
96 capacity (Goulson, 2010; Silva *et al.*, 2020). *L. malachurum* is a small, ground-nesting
97 eusocial species (Wyman and Richards, 2003). All three species have broad diets,
98 placing them in the top 10% most generalized bee species globally (Wood *et al.*, 2023).

99 Between February and May 2019, 252 female wild bees ($N_{A.flavipes} = 88$, $N_{B.terrestris} = 86$, and
100 $N_{L.malachurum} = 78$) were collected from 77 sampling sites across 15 locations (Figure 1;
101 Table S1). Sampling covered both SSF and LoA landscapes, the latter including semi-
102 natural areas such as shrublands, forests, nature reserves, and unmanaged meadows.
103 To reduce the likelihood of sampling closely related individuals, specimens within each
104 location were collected at least 300 meters apart.

105 Individuals were captured with sweeping nets, placed in individual 2 mL plastic tubes
106 filled with absolute ethanol, and stored at -80°C until molecular processing. Species
107 identity was confirmed through mitochondrial cytochrome oxidase I (COI) barcoding
108 prior to sequencing.

109 **2.2 | Land-use classification and landscape metrics**

110 Land-use classification was based on the Sentinel-2 10 m Land-Use/Land Cover (LULC)
111 Time Series dataset developed by ESRI and Impact Observatory (Karra *et al.*, 2021).
112 GeoTIFF files covering mainland Portugal were extracted and merged using *GDAL*
113 v.3.10.2 (Rouault *et al.*, 2025) and raster processing was conducted in *R* v.4.4.2 (R Core
114 Team, 2021) using the *terra* package v.1.8-5 (Hijmans, 2022). For each sampling site
115 (Table S1), a circular buffer with a 1 km radius was generated to characterize landscape
116 composition within the approximate foraging range of the three studied species
117 (Gathmann and Tschardt, 2002). Sampling sites were grouped at the location level,
118 resulting in 15 sampling locations (Table S1). Land-use proportions were calculated
119 after excluding water, flooded vegetation, snow/ice, and clouds (Table S2).

120 Principal component analysis (PCA) of centered and scaled land-use proportions was
121 used, together with geographic proximity, to assign the sampling locations to three
122 landscape categories: Small-Scale Farming landscape (SSF; $n = 6$; > 13 km separating
123 them), Near Low-Agricultural landscapes (Near_LoA; $n = 5$; 11-30 km from the nearest
124 SSF landscape), and Far Low-Agricultural landscapes (Far_LoA; $n = 4$; > 60 km from the
125 nearest SSF landscape) (Figure 1). Differences in land-use composition between SSF
126 and LoA landscapes were evaluated using Shapiro-Wilk tests for normality followed by
127 Wilcoxon rank-sum tests (Shapiro and Wilk, 1965; Haynes, 2013).

128 Landscape configuration was quantified using the *landscapemetrics* R package v.2.1.3
129 (Hesselbarth *et al.*, 2019). Twenty-three metrics describing fragmentation, edge
130 complexity, aggregation, patch size, diversity, and evenness were calculated and
131 summarized at the location level (Table S3-S4). Pairwise correlations were inspected

132 to assess redundancy, and PCA of centered and scaled landscape metrics was used to
133 derive four composite axes for downstream genotype-environment association
134 analyses (see Section 3.2).

135 **2.3 | DNA extraction and COI amplification**

136 Genomic DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen), following
137 the manufacturer's protocol with minor modifications. DNA extractions used thoracic
138 tissue and the fore and middle legs to reduce contamination from pollen, gut contents,
139 parasites, and endosymbionts. DNA quantity was measured using a Qubit 2.0
140 fluorometer (Thermo Scientific, Waltham, MA, USA), and DNA integrity was assessed
141 by 1% agarose gel electrophoresis.

142 Species identity was confirmed through COI barcoding (Macrogen Europe). Universal
143 primer pairs LepF/LepR (Hajibabaei *et al.*, 2006) and LCO1490/HCO2198 (Folmer *et al.*,
144 1994) successfully amplified COI fragments for *Bombus terrestris* and *Lasioglossum*
145 *malachurum*. However, amplification in *Andrena flavipes* was inconsistent using
146 universal primers. To address this, available *A. flavipes* COI sequences were manually
147 curated in *Sequencher* v.4.0.5 (Gene Codes Corporation) and aligned using *MAFFT*
148 v.7.271 (Kato and Standley, 2013) through *AliView* v.1.26. These alignments were used
149 to design species-specific primers (Table S5) with *Primer3* v.0.4.0. (Untergasser *et al.*,
150 2012). Following barcoding confirmation, DNA samples were shipped to CD Genomics
151 (Shirley, NY, USA) for RAD sequencing.

152 **2.4 | RADseq data processing and SNP calling**

153 Library preparation and RAD sequencing were outsourced to CD Genomics (Shirley, NY,
154 USA). A single-enzyme RADseq protocol using *Pst*I was applied, and paired-end
155 sequencing was performed on an Illumina NovaSeq platform using 2 x 150 bp reads.

156 Raw-read quality was assessed with *FastQC* v.0.12.0 (Andrews, S., 2010), and
157 summary reports were compiled using *MultiQC* v.1.28 (Ewels *et al.*, 2016). Adapter
158 trimming and quality filtering were performed with *fastp* v.0.23.4 (Chen, 2023). Reads
159 were trimmed to 140 bp, and only read pairs meeting the minimum length and quality
160 requirements were retained. Filtered FASTQ files were processed using
161 *process_radtags* from *Stacks* v.2.65 (Catchen *et al.*, 2013). Filtered reads were aligned
162 to the corresponding species-specific reference genome using *bwa-mem2* v.2.2.1. (Li,
163 2013) and *SAMtools* v.1.19.2. was used to process BAM files (Li *et al.*, 2009).

164 SNP calling was performed independently for each species using *Stacks ref_map.pl*
165 v.2.65 (Catchen *et al.*, 2013), requiring loci to be present in at least 80% of populations.
166 Pairwise relatedness was estimated using *VCFTools* v.0.1.16 (Manichaikul *et al.*, 2010;
167 Danecek *et al.*, 2011). Sample pairs with values > 0.20 were considered probable close
168 relatives and the individual (from the pair) with the higher proportion of missing data
169 was removed from downstream analyses. Genotype-level filtering was performed with
170 *BCFtools* v.1.18 (Li, 2011), masking genotypes with genotype quality < 20 or genotype
171 depth < 8 (Table S6).

172 Final filtering was conducted using *tidypopgen* R package v.0.4.3 (Carter *et al.*, 2026),
173 retaining biallelic SNPs with per-locus missingness < 20% and minor allele frequency
174 (MAF) > 0.05 (see Appendix I).

175 **2.5 | Population structure and genetic estimates**

176 Population structure and genetic estimations were assessed for each species using
177 the filtered SNP datasets. Because candidate outliers might represent a small fraction
178 of the full SNP datasets, their inclusion was not expected to substantially affect
179 estimates of population structure or genetic diversity.

180 Population structure was inferred using PCA (Zheng *et al.*, 2012), fineRADstructure
181 (Malinsky *et al.*, 2018), and ALStructure (Cabreros and Storey, 2019). Genetic variation
182 was characterized using observed (H_o) and expected (H_e) heterozygosity, nucleotide
183 diversity (π), the number of segregating sites, private alleles, standardized individual
184 heterozygosity (H_s), the proportion of polymorphic loci, the inbreeding coefficient (F_{IS}),
185 and global and pairwise genetic differentiation (F_{ST}). Isolation by distance was tested
186 using Mantel tests between log-transformed geographic distances among sampling
187 locations and linearized genetic distances, $F_{ST}/(1-F_{ST})$ (Rousset, 1997). All genetic
188 estimations and mantel tests were computed using the *snpr* package v.1.2.11
189 (Hemstrom and Jones, 2023) in R. Only locations represented by at least three
190 individuals were included in the analyses. Full diversity and differentiation estimates
191 are provided in Tables S7-S9.

192 **2.6 | Detection of putative loci under selection**

193 To identify loci displaying patterns consistent with putative selection, we implemented
194 a complementary set of genome scan and genotype-environment association
195 approaches. Selection detection methods included PCAdapt, BayeScan and BayPass
196 contrast (C2) analyses, whereas BayPass Bayes Factor (BF) analyses was used to
197 detect associations between allele frequencies and landscape variables. In addition,

198 BayPass local score analyses were applied to the outputs generated from BayPass
 199 contrast (C2) and Bayes Factor (BF).

200 This framework reduced reliance on a single statistical model and prioritized loci
 201 supported across complementary approaches. Method-specific candidates were
 202 retained in the results tables but not used for biological interpretation. All outlier scans
 203 were conducted independently for each species using the filtered SNPs datasets from
 204 Section 2.4.

205 PCAdapt analyses were conducted in R using the *pcadapt* R package v.4.4.1 (Privé *et*
 206 *al.*, 2020). After exploring K values using scree plots, final outlier scans were performed
 207 with K = 1, and p-values were corrected using the Benjamini-Hochberg false discovery
 208 rate (Tables S10-S12). This correction was selected for the final candidate set because
 209 it provided a conservative but not overly restrictive procedure. *BayeScan* v.2.1 was run
 210 using sampling location as the grouping factor, prior odds of 100 for the neutral model,
 211 a burn-in of 50,000 iterations, and a final sample size of 5,000 iterations (Foll and
 212 Gaggiotti, 2008) (Tables S13-S14).

213 *BayPass* v.3.1 was used to test predefined land-use contrasts while accounting for
 214 shared population history (Gautier, 2015). Two BayPass models were implemented.
 215 First, the core model was used to estimate the population covariance matrix (Ω) and to
 216 perform contrasts analyses using the C2 statistic. Three contrasts were tested
 217 independently for each species: SSF x LoA, SSF x Near_LoA, and SSF x Far_LoA. C2
 218 significance thresholds were calibrated using pseudo-observed datasets simulated
 219 under the fitted neutral model. Second, the auxiliary covariate (AUX) model was used

220 to test associations between SNP allele frequencies and the predefined land-use
221 contrasts, with candidate loci retained using a threshold of $BF(dB) > 10$.

222 To identify broader genomic regions, local score analyses were applied to the results
223 generated by the BayPass C2 and BF analyses (Fariello *et al.*, 2017). This approach
224 aggregates signals across neighboring markers to detect genomic regions with
225 clustered moderate signals. We used a penalty parameter of $\xi = 1$, which provides a
226 relatively sensitive balance between signal accumulation and penalization. Markers
227 with an allele frequency below 0.20 or above 0.80 were excluded ($min.maf = 0.20$),
228 thereby limiting the influence of weakly informative, low-frequency variants that are
229 unlikely to show robust differentiation. Detailed BayPass commands and local-score
230 parametrization are provided in Appendix I. BayPass C2 outliers can be seen in Tables
231 S15-S17, BayPass BF outliers in Tables S18-S20, and BayPass local scores in Tables
232 S21-S23.

233 Candidate SNPs from the different outlier methods were merged by chromosome and
234 genomic position (Tables S24-S26). Method overlap was visualized using UpSet plots
235 generated in R with *ComplexUpset* R package v.1.3.3 (Lex *et al.*, 2014). Multi-method
236 candidates were defined as loci supported by at least three approaches for *Andrena*
237 *flavipes* and *Bombus terrestris*. For *Lasioglossum malachurum*, where BayeScan
238 identified no candidates, multi-methods candidates were defined as loci supported by
239 at least two approaches.

240 **2.7 | Detection of putative loci associated with landscape configuration**

241 Associations between SNPs and landscape configuration were tested using *BayPass*
242 v.3.1 under the auxiliary covariate model (AUX) (Gautier, 2015). The four landscape

metrics PCA axes were used as composite environmental predictors. Analyses were conducted independently for each species using covariance matrices estimated under the BayPass core model. The AUX model was run using five independent random seeds, and decisive SNP-covariate associations were retained when $BF(dB) > 20$ (Tables S27-S29). Because the predictors were PCA axes summarizing correlated landscape metrics, GEA results were interpreted as associations with composite gradients of landscape configuration rather than with individual landscape variables.

2.8 | Functional SNP annotation and candidate gene characterization

Candidate SNPs were annotated independently for each species using the corresponding reference genome and the best available annotation resources (Tables S30-S32). Functional annotation was performed with *eggNOG-mapper* v.2.1.13 (Cantalapiedra *et al.*, 2021). For *Andrena flavipes* and *Bombus terrestris*, candidate SNPs were intersected species-specific annotation files to identify overlapping genes and intergenic regions. Intergenic SNPs were assigned to the nearest annotated gene within 50 kb, while retaining their intergenic classification. Because no native annotation file was available for *Lasioglossum malachurum*, gene structures were transferred from the *Lasioglossum baleicum* annotation to the *L. malachurum* reference genome using *Liftoff* v.1.6.3 (Shumate and Salzberg, 2021). *L. baleicum* was chosen since it provided the most complete and highest-quality annotation file available for the genus. Across all three species, only SNPs directly overlapping annotated genes were used for functional interpretation.

264 **3 | Results**

265 **3.1 | Sequencing summary and SNP calling**

266 A total of 252 samples were sequenced on an Illumina NovaSeq platform (2 x 150 bp).
267 Following initial quality control with FastQC, 35 samples were excluded because they
268 did not meet the required quality thresholds. These exclusions included four *Andrena*
269 *flavipes* samples, 18 *Bombus terrestris* samples, and 13 *Lasioglossum malachurum*
270 samples (Table S6).

271 Mean mapping rates averaged 96.9% (range: 89.9 - 99.8%) for *A. flavipes*, 93.4% (range:
272 88.2% - 98.1%) for *B. terrestris*, and 90.0% (range: 80.4% - 96.7%) for *L. malachurum*.
273 Mean per-sample coverage ranged between 22.7x and 27.3x across species. Kinship
274 analysis identified one pair of closely related individuals in *B. terrestris* (--
275 *relatedness2* > 0.20), from which the individual with higher missingness,
276 BOM_SAH_002_poll504, was removed. No individuals were removed due to
277 relatedness in *A. flavipes* or *L. malachurum*.

278 After quality control, removal of ambiguous samples, genotype masking, and final SNP
279 filtering, the datasets comprised 78 individuals and 129,124 SNPs for *A. flavipes*, 62
280 individuals and 52,479 SNPs for *B. terrestris*, and 59 individuals and 87,221 SNPs for *L.*
281 *malachurum*. Full sequencing, filtering, mapping, SNP retention parameters and
282 statistics are provided in Appendix I and Table S6.

283 **3.2 | Land-use categories and landscape metrics**

284 Land-use composition differed markedly between small-scale farming (SSF) and low-
285 agricultural (LoA) landscapes. LoA landscapes were dominated by semi-natural land-

286 use classes, mainly rangelands and tree cover, whereas SSF landscapes exhibited
287 higher proportions of croplands and built areas (Figure S1; Table S2). PCA of land-use
288 composition separated SSF locations from LoA locations, with PC1 capturing the main
289 contrast between cropland/built area and rangelands/tree cover (Figure 2; Figure S2).
290 Most SSF locations showed positive PC1 scores and most LoA locations showed
291 negative PC1 scores (Figure 2). Statistical comparisons supported these differences
292 between SSF and LoA landscapes and among SSF, Near_LoA, and Far_LoA (Figure S3-
293 S4). Assafora site deviated from the general SSF pattern and was therefore excluded
294 from subsequent analyses to maintain a clear separation between landscape
295 categories.

296 Landscape metric correlations showed redundancy among metrics describing
297 fragmentation, edge complexity, diversity, aggregation, and patch structure (Figure S5;
298 Table S3). The biplots of landscape metrics further differentiated SSF from LoA
299 landscapes (Figure S6). PC1 (fragmentation and compositional complexity; 53.8% of
300 variation) captured the main contrast between fragmented, edge-rich, compositionally
301 complex landscapes and more cohesive landscapes with larger and more aggregated
302 patches. Positive PC1 values were associated mainly with number of patches (np),
303 patch density (pd), total edge (te), edge density (ed), and conditional entropy (condent).
304 Negative PC1 values were associated with patch cohesion (cohesion), standard
305 deviation of patch area (area_SD), largest patch index (lpi), and the degree to which
306 land-use classes are aggregated (contag) (Figure 2).

307 PC2 (patch heterogeneity and edge complexity; 15.5% of variation) primarily reflected
308 a contrast between patch size heterogeneity and cohesion on one side, and

309 edge/configurational complexity on the other (Figure S6). Positive PC2 values were
310 associated with area_sd, cohesion, Shannon's diversity index (shdi), and Simpson's
311 diversity index (sidi), whereas negative PC2 values were associated with number of
312 patches (np), patch density (pd), total edge (te), edge density (ed) and conditional
313 entropy (condent) (Figure 2).

314 PC3 (landscape diversity and configurational complexity; 12.0% of variation) mainly
315 represented variation in landscape diversity and configurational complexity (Figure 2;
316 Figure S6). Positive PC3 values were associated with Shannon's diversity index (shdi),
317 entropy (ent), and joint entropy (jointent). Negative PC3 values were associated with
318 cohesion, area_SD, lpi, and contag.

319 PC4 (patch density and connectivity variation; 10.4% of variation) captured a
320 secondary configuration gradient contrasting shdi, jointent, contag, and lpi with np, pd,
321 cohesion, and area_SD (Figure 2; Figure S6).

322 Taken together, these results indicate that SSF landscapes were more subdivided,
323 edge-rich, and configurationally complex, whereas LoA landscapes showed greater
324 cohesion, larger and more dominant patches, stronger aggregation, and greater
325 internal patch contiguity (Figure 2; Figure S2 and Figure S6). The four retained landscape
326 metrics PCA axes were used as composite predictors in GEA analysis.

327 **3.3 | Genetic diversity and population structure**

328 Genetic diversity was similar between SSF and LoA landscapes in all three species
329 (Tables S7-S9). Among species, *Lasioglossum malachurum* exhibited the highest levels
330 of genetic diversity ($H_o = 0.17$; $H_e = 0.17$), followed by *Bombus terrestris* ($H_o = 0.14$; H_e

331 = 0.15), whereas *Andrena flavipes* showed the lowest diversity ($H_o = 0.08$; $H_e = 0.10$).
 332 Inbreeding coefficients (F_{IS}) were low across species, with no evidence of strong
 333 inbreeding within sampling locations. Pairwise F_{ST} values were generally close to zero,
 334 with no clear differentiation by sampling location, landscape category, or geographic
 335 proximity (Figures S7-S9). SNP-based PCA revealed extensive overlap among
 336 individuals from different sampling locations and no clear clustering by landscape
 337 category in any species (Figure S10). fineRADstructure and ALStructure showed similar
 338 patterns, with weak population structure and no consistent separation among SSF,
 339 Near_LoA, and Far_LoA landscapes (Figures S11-S16). Mantel tests provided no
 340 evidence of isolation by distance in *Andrena flavipes* ($r = -0.52$, $p = 0.98$), *Bombus*
 341 *terrestris* ($r = 0.19$, $p = 0.26$), or *Lasioglossum malachurum* ($r = -0.23$, $p = 0.79$) (Figures
 342 S17-S19). Together, these results indicate that genetic diversity was maintained across
 343 SSF and LoA landscapes, while population structure remained weak in all species.

344 **3.4 | Detection of candidate loci under selection**

345 Outlier methods differed substantially in the number of candidate SNPs recovered, with
 346 PCAdapt identifying the largest candidate sets and BayeScan the smallest (Tables S10-
 347 S14; Figures S20-S21). BayPass C2, BayPass BF, and local score analyses provided
 348 complementary evidence for contrast-associated differentiation across SSF x LoA, SSF
 349 x Near_LoA, and SSF x Far_LoA contrasts (Tables S15-S23). These analyses revealed
 350 species-specific differences in the strongest contrasts (Figure 3; Figures S22-S33).

351 In *Andrena flavipes*, BayPass C2, BF, and local score analyses were concordant in
 352 identifying SSF x Far_LoA as the contrast displaying higher number of putatively
 353 selected SNPs. This contrast had the highest number of C2 candidates (2,267), the

354 highest number of BF candidates (129), and the strongest local score support, with 10
355 significant windows comprising 282 SNPs (Figure 4a; Tables S15, S18, S21; Figures S22,
356 S25-S27).

357 In *Bombus terrestris*, C2 and local score analyses both pointed to SSF x Near_LoA as
358 the strongest contrast. This contrast had the highest number of C2 candidates (1,826)
359 and the strongest local score support, with 40 significant windows comprising 1,083
360 SNPs (Figure 4b), whereas BF candidates were similar across contrasts (Tables S16,
361 S19, S22; Figures S23, S28-S30). This pattern suggests that, in *B. terrestris*, the
362 strongest evidence for contrast-associated differentiation came from the comparison
363 between SSF landscapes and Near_LoA landscapes.

364 In *Lasioglossum malachurum*, BayPass C2 and BF analyses identified SSF x LoA as the
365 strongest contrast, with 1,041 C2 candidates and 65 BF candidates (Tables S17, S21;
366 Figure S24). Local score analyses helped clarify this broad pattern by showing that
367 clustered signals were not restricted to the overall SSF x LoA contrast. Although SSF x
368 LoA had the highest number of significant local score windows, with three windows
369 (Figure 4c), SSF x Far_LoA contained the highest number of SNPs within local score
370 windows, with 65 SNPs across two windows (Table S23; Figures S31-S33). This suggests
371 that the broad SSF x LoA signal detected by single-SNP methods may partly reflect a
372 stronger contribution of the SSF x Far_LoA contrast.

373 After integrating support across methods and land use contrasts, 46 candidate SNPs in
374 *A. flavipes*, 60 in *B. terrestris*, and 48 in *L. malachurum* were retained as conservative
375 candidate sets for downstream interpretation (Figure 3a-c; Tables S24-S26). The
376 overlap among land-use contrasts revealed that most candidate SNPs were contrast-

377 specific, with relatively few candidates shared among contrasts (Figure 3d-f). Overall,
378 the results from the outlier framework suggest that candidate genomic differentiation
379 was limited in magnitude but non random across species and contrasts.

380 **3.5 | Genotype-environment association analyses**

381 Genotype-environment association analyses detected decisive SNP-covariate
382 associations in all three species, with species-specific distributions across landscape
383 PCA axes (Figure 3g; Tables S27-S29). In *Andrena flavipes*, 1,608 SNP-covariate
384 associations were classified as decisive ($BF(dB) > 20$) (14.0%). The decisive
385 associations were distributed relatively evenly across the four axes: PC1 accounted for
386 456 associations (28.4%), PC2 for 391 (24.3%), PC3 for 399 (24.8%), and PC4 for 362
387 (22.5%) (Figure 3g; Table S27). A similar pattern was observed in *Bombus terrestris*,
388 where 916 SNP-covariate associations were classified as decisive (14.8%): 261 with
389 PC1 (28.5%), 201 with PC2 (21.9%), 243 with PC3 (26.5%), and 211 with PC4 (23.0%)
390 (Figure 3g; Table S28). These distributions suggest that several dimensions of
391 landscape configuration contributed to allele-frequency variation in both species.

392 *Lasioglossum malachurum* showed the highest number and proportion of decisive
393 associations, with 2,062 SNP-covariate associations classified as decisive (16.1%)
394 (Figure 3g; Table S29). In this species, decisive associations were more concentrated
395 on PC3, which accounted for 768 associations (37.2%), compared with 424 for PC1
396 (20.6%), 475 for PC2 (23.0%), and 395 for PC4 (19.2%). This pattern suggests a stronger
397 association between allele-frequency variation and the landscape gradient related to
398 patch contiguity, internal patch structure, and patch richness.

399 **3.6 | Functional annotation of candidate SNPs**

400 Functional annotation of the conservative candidate SNP sets showed that several
401 candidate loci directly overlapped annotated gene models (Tables S30-S32). In
402 *Andrena flavipes*, 27 candidate SNPs overlapped annotated genes, corresponding to 24
403 unique genes (Table S30). In *Bombus terrestris*, 48 candidate SNPs overlapped
404 annotated genes, corresponding to 29 unique genes (Table S31). In *Lasioglossum*
405 *malachurum*, 34 candidate SNPs overlapped gene models, corresponding to 28 unique
406 genes (Table S32).

407 Across species, candidate genes were repeatedly associated with signal transduction,
408 cell adhesion, cytoskeletal organization, transcriptional regulation, ion transport, and
409 developmental processes (Figures S34-S36). Several annotations pointed to cell-cell
410 communication and tissue organization, including cadherin/protocadherin-related
411 genes in *A. flavipes* (*CDH17*, *DCHS2*, *FAT1*, *PCDH15*; Figure S34), immunoglobulin-
412 domain and extracellular matrix genes in *B. terrestris* (*igSF C2 domain*, *nrm*, *HSPG2*;
413 Figure S35), and immunoglobulin-domain, receptor, and scaffold genes in *L.*
414 *malachurum* (*Dscam1*, *Lar/PTPRD*, *Dlg1-like*; Figure S36).

415 Developmental and transcriptional regulation was also represented in all three
416 species. In *A. flavipes*, candidate genes included transcription-related annotations
417 such as *GB43937-like*, *cwo-like*, and *RBL2*, together with genes linked to planar cell
418 polarity and developmental patterning (*DCHS2*, *FAT1*, *EXT2*). In *B. terrestris*, candidate
419 genes included *MED20*, *SOX6*, and *HOXA2*, the latter associated with gnathal and
420 mouthpart patterning. In *L. malachurum*, candidate genes included
421 *HOXA2/proboscipedia-like*, *Soxn/SOX2-like*, *Klf12-like*, *rotund/ZNF362-like*, *NOTCH1-*

422 *like*, and *Mastermind-like*, indicating transcriptional and developmental regulatory
423 functions, including appendage or mouthpart identity.

424 Other recurrent functional themes included ion transport and neural or sensory-
425 related processes, represented by *GRIK1* in *A. flavipes*, *cac* and *UNC13c* in *B. terrestris*,
426 and *Odorant receptor 4-like*, *Dscam1*, and *Lar/PTPRD* in *L. malachurum*. Genes
427 associated with protein modification, metabolism, and extracellular signaling were
428 also present, including *METAP2*, *AMPD3*, and *NOX4* in *A. flavipes*, *ATG16L1* and acetyl-
429 CoA carboxylase-like annotations in *B. terrestris*, and *Mib1*, *Sulf1*, *Hs3st-a/HS3ST5-*
430 *like*, and *Fur2-like* in *L. malachurum*.

431 Cross-referencing the functionally annotated conservative candidate SNPs with the
432 BayPass GEA decisive associations (BF(dB) > 20) results identified two overlapping
433 SNPs in *A. flavipes*, four in *B. terrestris*, and three in *L. malachurum*, all associated with
434 PC4 (i.e., secondary configuration gradient). In *A. flavipes*, these decisive SNPs
435 overlapped *AMPD3*, *EXT2*, and *RBL2*; in *B. terrestris*, they overlapped *TIAL1*, *MED20*,
436 *SOX6*, a sugar transporter-like gene, and one gene without functional annotation ; and
437 in *L. malachurum*, they overlapped *Mesr6/C5orf22-like*, a chitin-binding domain, and a
438 *TNS3-like* gene.

439 Overall, candidate genes were associated with biological processes potentially
440 relevant to responses to heterogeneous small-scale farming landscapes.

441 **4 | Discussion**

442 This study assessed how SSF landscapes are associated with genetic diversity,
443 population structure, and putative adaptive genomic differentiation in three wild bee

species. Across the three studied species, genetic diversity was similar, genetic differentiation was low, population structure was weak, and isolation by distance was not detected, indicating high connectivity among sampling locations despite marked differences in landscape composition and configuration. At the same time, the outlier framework and GEA analyses identified conservative candidate SNP sets and species-specific genomic associations with landscape configuration.

Together, these findings suggest that allele-frequency shifts can occur in human-modified landscapes even when genome-wide differentiation remains low (Theodorou *et al.*, 2018, 2021; Jackson *et al.*, 2018, 2020; Kardum Hjort *et al.*, 2025). However, candidate SNPs were not uniform across species, contrasts, or outlier methods. This interspecific heterogeneity suggests that the same landscape does not lead to a similar genomic response across different wild bee taxa. Instead, genomic responses to landscape heterogeneity are likely mediated by species-specific ecological traits, such as, dispersal capacity, sociality, and the degree to which each species can exploit resources in these SSF landscapes.

4.1 | High gene flow maintains genetic diversity and weak population structure

Genetic diversity was similar across SSF and LoA landscapes in all three species, with no consistent reduction in diversity associated with either landscape category. Inbreeding coefficients were positive, pairwise F_{ST} values were close to zero, and PCA, fineRADstructure, and ALStructure revealed no clear clustering by landscape category. These patterns are consistent with high connectivity among sampling locations and with previous genomic studies of common and widespread wild bees, which frequently

467 report weak population structure across heterogeneous and human-modified
468 landscapes (Theodorou *et al.*, 2018, 2021; Jackson *et al.*, 2018, 2020; Silva *et al.*, 2020;
469 Kelemen and Rehan, 2021; Hart *et al.*, 2022; Sousa *et al.*, 2023, 2025; Kardum Hjort *et*
470 *al.*, 2025).

471 High connectivity in widespread wild bees may reflect dispersal capacity, large
472 effective population sizes, and demographic stability over time (Theodorou *et al.*, 2018;
473 Silva *et al.*, 2020). Although human-modified environments can restrict movement and
474 reduce gene flow in some pollinators, the SSF landscape studied here retains semi-
475 natural elements, heterogeneous land-use mosaics, and floral resources. Under these
476 conditions, SSF landscapes may function less as barriers to dispersal and more as
477 permeable matrices that support connectivity among wild bee populations. However,
478 this interpretation is likely to be species-specific. Bees with limited dispersal ability,
479 narrower habitat or floral specialization, specific nesting requirements may perceive
480 the same landscape as fragmented, resulting in lower connectivity and stronger
481 population genetic structure.

482 **4.2 | Genomic differentiation associated with small-scale farming** 483 **landscapes**

484 Despite high gene flow and limited genetic differentiation, the outlier framework
485 identified a conservative subset of SNPs associated with contrasts between SSF and
486 LoA landscapes. These candidates were supported by complementary approaches,
487 including PCAdapt, BayeScan, BayPass C2 and BF statistics, and local score analyses.
488 The small conservative candidate sets likely reflect extensive gene flow, multi method

489 support requirements, and reduced genomic representation (François *et al.*, 2016;
490 Ahrens *et al.*, 2018).

491 Although high gene flow tends to homogenize genomic variation, it does not necessarily
492 prevent adaptive divergence. Selection may act on specific loci while the rest of the
493 genome remains homogenized by migration, generating detectable allele-frequency
494 shifts associated with landscape heterogeneity (Kawecki and Ebert, 2004). At the same
495 time, extensive connectivity may constrain the magnitude of local adaptation, favoring
496 scenarios in which adaptive responses may be distributed across many loci of small
497 effect rather than concentrated in highly differentiated genomic regions (Barghi *et al.*,
498 2020; Yeaman, 2022). Such patterns have been reported in other widespread and
499 generalist pollinators, where signatures of environmental association or putative
500 adaptation occur despite high gene flow and weak population structure (Theodorou *et al.*,
501 2018; Jackson *et al.*, 2018, 2020; Colgan *et al.*, 2022; Hart *et al.*, 2022; Heraghty *et al.*,
502 2023; Kardum Hjort *et al.*, 2024, 2025).

503 Local score analyses added an important layer to the single-SNP results by identifying
504 whether differentiation signals were clustered in genomic windows (Fariello *et al.*,
505 2017). Among the three species, clustering was most pronounced in *Andrena flavipes*.
506 One possible explanation is that the genus *Andrena* is one of the most species-rich bee
507 genera globally and has undergone substantial diversification, particularly in the
508 Western Palearctic and Mediterranean basin (Bossert *et al.*, 2022). As seen in Herrera
509 *et al.*, (2025), the genus *Andrena* represents a relatively young and rapidly radiating
510 monophyletic lineage, with multiple cases of recent niche differentiation and
511 speciation. Such rapid diversification may leave standing genetic variation or

512 structured genomic regions on which selection can act, increasing the number of
513 significant local score windows.

514 In *Bombus terrestris*, local score clustering was stronger for the SSF x Near_LoA
515 contrast. As a highly mobile and generalist social species, *B. terrestris* is expected to
516 experience extensive gene flow across landscapes, which can limit broad genomic
517 differentiation while still allowing for localized allele-frequency shifts at specific loci
518 under selection. In addition, its ecological flexibility and tolerance of human-modified
519 environments have contributed to its widespread distribution and persistence across
520 Europe (Owen *et al.*, 2013; Martinet *et al.*, 2015; Theodorou *et al.*, 2018; Herbertsson *et*
521 *al.*, 2021). Thus, the restricted local score clustering is consistent with selection acting
522 against extensive gene flow in a highly mobile species.

523 *Lasioglossum malachurum* showed fewer local score windows than *A. flavipes*,
524 although clustered signals were still detected across the three contrasts (Figure 19-21).
525 This more limited clustering may be consistent with more diffuse or polygenic
526 responses, but it may also reflect species-specific life-history traits. *L. malachurum* is
527 a common polylectic ground nesting species frequently associated with disturbed or
528 human-managed habitats, and its mating system, including facultative polyandry and
529 off-nest mating, may promote local admixture and reduce the formation of strongly
530 differentiated genomic regions (Wyman and Richards, 2003; Weissel *et al.*, 2006; Sousa
531 *et al.*, 2023). Together, these traits may maintain gene flow and limit broad genomic
532 differentiation while still allowing localized responses to selection.

533 Overall, the results suggest that the genomic architecture of putative adaptive
534 responses to SSF landscapes may vary among species and contrasts. Although all

535 three species are widespread, polylectic, and abundant in SSF and LoA landscapes,
536 they differ markedly in body size, social organization, dispersal capacity, and ecological
537 requirements, although the present data cannot disentangle these mechanisms
538 directly.

539 **4.3 | Landscape configuration as a driver of genotype-environment** 540 **associations**

541 Genotype-environment association analyses showed that decisive associations were
542 distributed across multiple axes of landscape configuration rather than being restricted
543 to a single landscape gradient. This indicates that allele-frequency variation was
544 associated not only with broad differences between SSF and LoA landscapes, but also
545 with finer scale aspects of landscape structure. Beyond influencing the spatial and
546 temporal continuity of floral resources (Hass *et al.*, 2018; Sirami *et al.*, 2019), landscape
547 configuration may also affect the composition of bee associated microbial
548 communities (Cariveau *et al.*, 2014), parasite and pathogen dynamics (McNeil *et al.*,
549 2020; Cohen *et al.*, 2021), and other biotic interactions. Moreover, landscape
550 configuration in agricultural systems is also correlated with management intensity,
551 including pesticide use, fertilizer application, and soil disturbance. Consequently,
552 these gradients likely capture multiple ecological variables that together shape the
553 adaptive responses.

554 The distribution of decisive GEA associations differed among species. In *Andrena*
555 *flavipes* and *Bombus terrestris*, associations were relatively evenly distributed across
556 the four landscape PCA axes, with PC1 (fragmentation and compositional complexity)
557 contributing slightly more than the remaining axes. This pattern suggests that allele-

558 frequency variation in these species was associated with multiple, rather than a single
559 dominant, dimension of landscape configuration. In contrast, *Lasioglossum*
560 *malachurum* showed both the highest total number of decisive associations and a
561 greater contribution from PC3 (patch contiguity, internal structure, and patch richness).
562 This stronger association may reflect species-specific ecological traits. As the smallest
563 of the three study species, *L. malachurum* is expected to have shorter foraging and
564 dispersal distances than *A. flavipes* and especially *B. terrestris*, as bee foraging range
565 scales positively with body size (Gathmann and Tscharntke, 2002; Greenleaf *et al.*,
566 2007). Species traits such as body size, dispersal capacity, and sociality are also known
567 to modify sensitivity to habitat loss and fragmentation (Bommarco *et al.*, 2010), making
568 smaller species potentially more responsive to variation in patch contiguity and internal
569 habitat structure.

570 These findings support the hypothesis that loci associated with landscape
571 configuration can be detected even in the presence of high gene flow, consistent with
572 other landscape genomic studies of wild bees (Theodorou *et al.*, 2018; Jackson *et al.*,
573 2018; Kelemen and Rehan, 2021; Colgan *et al.*, 2022; Hart *et al.*, 2022). The species-
574 specific distribution of associated loci further suggests that responses to small-scale
575 farming landscapes are multifaceted, driven by several landscape composition
576 variables, and shaped by ecological traits and sensitivity to landscape structure.
577 However, these associations should not be interpreted as evidence that individual
578 landscape variables are causal selective agents. Instead, GEA identifies composite
579 landscape gradients associated with allele-frequency variation, requiring future
580 ecological and functional validation.

581 **4.4 | Functional interpretation of conservative candidate loci**

582 Functional annotation provided biological context but should be interpreted cautiously.
583 Across species, candidate genes were associated with signaling, cell adhesion,
584 cytoskeletal organization, transcriptional regulation, ion transport, sensory or neural
585 processes, metabolism, and protein modification. These broad categories are
586 compatible with biological processes that may mediate responses to heterogeneous
587 environments, including sensory perception, development, behavior, stress response,
588 immunity, and physiological regulation (Theodorou *et al.*, 2018; Jackson *et al.*, 2018,
589 2020; Henriques *et al.*, 2018; Jaffé *et al.*, 2019; Kelemen and Rehan, 2021).

590 The subset of functionally annotated conservative candidates that also showed decisive
591 genotype-environment associations provides an additional link between candidate
592 gene function and landscape configuration. Notably, all nine decisive overlapping SNPs
593 were associated with PC4, despite decisive GEA associations being more generally
594 distributed across all four landscape configuration axes. These SNPs overlapped genes
595 associated with metabolism, transcriptional and developmental regulation,
596 extracellular organization, membrane transport, and chitin-related processes.

597 Overall, functional annotation identified functional categories consistent with previous
598 studies. Although the specific biological roles differed among species, several
599 overlapping genes were related to cell communication, tissue organization,
600 transcriptional or developmental regulation, and sensory or structural processes. This
601 pattern is consistent with the species-specific GEA results and suggests that
602 landscape configuration may be associated with different functional components of
603 genomic variation in each species. However, these annotations indicate only that

604 candidate SNPs occur within or near genes with these predicted functions, providing
605 evidence to complement outlier and genotype-environment association analyses.
606 Because landscape configuration may be correlated with agrochemical exposure and
607 management practices, these candidate genes may also reflect adaptive responses to
608 correlated selective pressures rather than landscape configuration alone.

609 **5 | Final considerations**

610 This study found a consistent pattern of genetic diversity, weak population structure,
611 absence of isolation by distance, and extensive connectivity across contrasting
612 landscapes. At the same time, the outlier framework identified conservative candidate
613 SNP sets associated with land-use contrasts and gradients of landscape configuration.
614 These signals were limited in genomic extent, but their repeated support across
615 complementary analyses suggests that SSF landscapes may be associated with
616 detectable allele-frequency shifts. Importantly, the distribution of candidate loci and
617 GEA signals differed among species, indicating that genomic responses to agricultural
618 landscape heterogeneity are taxon-specific rather than uniform.

619 However, some caveats should be considered when interpreting the candidate loci
620 identified in this study. First, the conservative outlier framework reduced the risk of
621 false positives but may also exclude weaker, or polygenic responses (François *et al.*,
622 2016; Ahrens *et al.*, 2018). Candidate loci supported by a single method should not be
623 treated as false positives, but they require more cautious interpretation than multi-
624 method candidates.

625 Second, although multiple locations were sampled per land-use category, sample
626 sizes per location were small and uneven among locations and species. Large SNP

627 datasets can improve estimates of population genetic parameters and increase power
628 for outlier detection, but they do not fully compensate for unbalanced sampling.

629 Third, RAD sequencing captures only a fraction of the genome and may miss loci or
630 regions under selection. Whole-genome resequencing would provide finer resolution
631 and improve evaluation of genomic architecture in these species. Fourth, GEA analyses
632 cannot identify the precise selective agents when land-use predictors are PCA axes.

633 Finally, functional annotation remains provisional: candidate SNPs overlapping genes
634 do not demonstrate functional effects and nearest-gene assignments cannot be
635 assumed to indicate regulation. Future work should test whether these candidate SNPs
636 and genomic regions also contribute directly to adaptive responses in other wild bee
637 species.

638 Despite these caveats, the results suggest that SSF landscape did not act as a strong
639 barrier to dispersal at the spatial scale sampled and that gene flow was maintained
640 despite marked differences in landscape composition and configuration.
641 Heterogeneous SSF landscapes may therefore maintain connectivity and standing
642 genetic variation while also being associated with detectable allele-frequency shifts.
643 The species-specific nature of the genomic responses observed here further indicates
644 that adaptation to SSF landscapes is unlikely to follow a single evolutionary trajectory,
645 emphasizing the importance of multi-species approaches when assessing wild bee
646 responses to land-use change.

647 These findings are particularly relevant in the context of ongoing and projected climate
648 and land-use change. Recent climate change projections have identified the Oeste
649 Region as one of the most important areas for biodiversity persistence and species

650 redistribution under future environmental scenarios, highlighting its potential role as a
651 refugia for multiple taxa (Araújo, 2022). Although we did not test future climatic
652 responses, the combination of genetic diversity, connectivity, and candidate adaptive
653 variation suggest that these landscapes may contribute to the evolutionary resilience
654 under future environmental change.

655 Importantly, the significance of conserving these landscapes extends beyond their role
656 as future biodiversity refugia. The west Region encompasses one of the most important
657 small-scale farming landscapes in Portugal, supporting a highly heterogeneous
658 agricultural mosaic that contributes simultaneously to food production, biodiversity
659 conservation, and cultural heritage. Preserving this landscape structure may therefore
660 support crop production, pollinator communities, ecosystem services, and traditional
661 agricultural systems with strong ecological and cultural value.

662 **Author contributions**

663 **AFH:** formal analysis, visualization, sampling, investigation, writing – original draft.
664 **PdS:** collected samples, formal analysis, visualization, investigation, writing – review
665 and editing. **OSP:** conceptualization, funding acquisition, project administration,
666 formal analysis, methodology, supervision, writing – review and editing. **LGC:**
667 conceptualization, funding acquisition, project administration, formal analysis,
668 methodology, supervision, writing – review and editing. **DG:** formal analysis, writing –
669 review and editing. **SES:** formal analysis, writing – review and editing. **OR:** collected
670 samples, writing – review and editing. **TJW:** collected samples, writing – review and
671 editing.

672 **Competing Interests**

673 The authors have no conflict of interest to declare.

674 **Data Archiving**

675 The raw sequence data used for this study will be deposited on NCBI after acceptance
676 for publication. Other data and code used for this study can be found in the Supporting
677 Information and Appendix I, and could be available, upon request to the authors.

678 **Ethical Statement**

679 Permits for the collection of wild bee specimens were issued by the Instituto da
680 Conservação da Natureza e das Florestas (ICNF) under license no. 486/2019/CAPT.

681 **References**

Ahrens CW, Rymer PD, Stow A, Bragg J, Dillon S, Umbers KDL, *et al.* (2018). The search for loci under selection: trends, biases and progress. *Mol Ecol* **27**: 1342–1356.

Albrecht M, Bossart S, Tschanz P, Keller T, Sutter L (2023). Grassland extensification enhances nest densities of ground-nesting wild bees. *J Appl Ecol* **60**: 2550–2560.

Andrews, S. (2010). Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data.

Araújo MB (2022). *Biodiversidade 2030: Nova Agenda Para a Conservação Em Contexto De Alterações Climáticas*.

Baldock DW, Wood TJ, Cross I, Smit J (2018). *The Bees of Portugal (Hymenoptera: Apoidea: Anthophila)*.

Barghi N, Hermisson J, Schlötterer C (2020). Polygenic adaptation: a unifying framework to understand positive selection. *Nat Rev Genet* 2020 2112 **21**: 769–781.

Bommarco R, Biesmeijer JC, Meyer B, Potts SG, Pöry J, Roberts SPM, *et al.* (2010). Dispersal capacity and diet breadth modify the response of wild bees to habitat loss. *Proc R Soc B Biol Sci* **277**: 2075–2082.

Bossert S, Wood TJ, Patiny S, Michez D, Almeida EAB, Minckley RL, *et al.* (2022). Phylogeny, biogeography and diversification of the mining bee family Andrenidae. *Syst Entomol* **47**: 283–302.

Brandt K, Glemnitz M, Schröder B (2017). The impact of crop parameters and surrounding habitats on different pollinator group abundance on agricultural fields. *Agric Ecosyst Environ* **243**: 55–66.

Cabreros I, Storey JD (2019). A Likelihood-Free Estimator of Population Structure Bridging Admixture Models and Principal Components Analysis. *Genetics* **212**: 1009–1029.

Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol Biol Evol* **38**: 5825–5829.

Cariveau DP, Elijah Powell J, Koch H, Winfree R, Moran NA (2014). Variation in gut microbial communities and its association with pathogen infection in wild bumble bees (*Bombus*). *ISME J* **8**: 2369–2379.

Carter EJ, Tysall EE, Hodgson JA, Manica A (2026). tidyPopgen: Tidy population genetics in R. *Methods Ecol Evol* **17**: 480–487.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013). Stacks: an analysis tool set for population genomics. *Mol Ecol* **22**: 3124–3140.

Chen S (2023). Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta* **2**: e107.

Cohen H, Smith GP, Sardiñas H, Zorn JF, McFrederick QS, Woodard SH, *et al.* (2021). Mass-flowering monoculture attracts bees, amplifying parasite prevalence. *Proc R Soc B Biol Sci* **288**: 20211369.

Colgan TJ, Arce AN, Gill RJ, Ramos Rodrigues A, Kante A, Duncan EJ, *et al.* (2022). Genomic Signatures of Recent Adaptation in a Wild Bumblebee (K Tamura, Ed.). *Mol Biol Evol* **39**: msab366.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, *et al.* (2011). The variant call format and VCFtools. *Bioinforma Oxf Engl* **27**: 2156–8.

Eurostat (2022). Farms and farmland in the European Union. *Farms Farml Eur Union*.

Ewels P, Magnusson M, Lundin S, Käller M (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**: 3047–3048.

Fariello MI, Boitard S, Mercier S, Robelin D, Faraut T, Arnould C, *et al.* (2017). Accounting for linkage disequilibrium in genome scans for selection without individual genotypes: The local score approach. *Mol Ecol* **26**: 3700–3714.

Finn C, Grattarola F, Pincheira-Donoso D (2023). More losers than winners: investigating Anthropocene defaunation through the diversity of population trends. *Biol Rev* **98**: 1732–1748.

Foll M, Gaggiotti O (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* **180**: 977–93.

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Biol Biotechnol* **3**: 294–9.

François O, Martins H, Caye K, Schoville SD (2016). Controlling false discoveries in genome scans for selection. *Mol Ecol* **25**: 454–469.

Gathmann A, Tscharnkte T (2002). Foraging ranges of solitary bees. *J Anim Ecol* **71**: 757–764.

Gautier M (2015). Genome-Wide Scan for Adaptive Divergence and Association with Population-Specific Covariates. *Genetics* **201**: 1555–1579.

Glück M, Geue JC, Thomassen HA (2022). Environmental differences explain subtle yet detectable genetic structure in a widespread pollinator. *BMC Ecol Evol* **22**: 8–8.

Goulson D (2010). *Bumblebees: behaviour, ecology, and conservation*. Oxford University Press on Demand.

Grab H, Branstetter MG, Amon N, Urban-Mead KR, Park MG, Gibbs J, *et al.* (2019). Agriculturally dominated landscapes reduce bee phylogenetic diversity and pollination services. *Science* **363**: 282–284.

Greenleaf SS, Williams NM, Winfree R, Kremen C (2007). Bee foraging ranges and their relationship to body size. *Oecologia* **153**: 589–596.

Gutiérrez-Briceño I, García-Llorente M, Ortega-Marcos JJ, Azcárate FM, Hevia V (2023). Exploring the effect of landscape composition and agroecological practices on wild bees in horticultural farms. *Basic Appl Ecol* **71**: 33–44.

Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PDN (2006). DNA barcodes distinguish species of tropical Lepidoptera. *Proc Natl Acad Sci* **103**: 968–971.

Hart AF, Verbeeck J, Ariza D, Cejas D, Ghisbain G, Honchar H, *et al.* (2022). Signals of adaptation to agricultural stress in the genomes of two European bumblebees. *Front Genet* **13**: 993416.

Hass AL, Kormann UG, Tschardt T, Clough Y, Baillod AB, Sirami C, *et al.* (2018). Landscape configurational heterogeneity by small-scale agriculture, not crop diversity, maintains pollinators and plant reproduction in western Europe. *Proc Biol Sci* **285**: 20172242–20172242.

Haugen H, Dervo BK, Østbye K, Heggenes J, Devineau O, Linløkken A (2024). Genetic diversity, gene flow, and landscape resistance in a pond-breeding amphibian in agricultural and natural forested landscapes in Norway. *Evol Appl* **17**: e13633.

Haynes W (2013). Wilcoxon Rank Sum Test. In: *Encyclopedia of Systems Biology*, Springer, New York, NY, pp 2354–2355.

Hemstrom W, Jones M (2023). snpR: User friendly population genomics for SNP data sets with categorical metadata. *Mol Ecol Resour* **23**: 962–973.

Henriques D, Wallberg A, Chávez-Galarza J, Johnston JS, Webster MT, Pinto MA (2018). Whole genome SNP-associated signatures of local adaptation in honeybees of the Iberian Peninsula. *Sci Rep* **8**: 11145.

Heraghty SD, Jackson JM, Lozier JD (2023). Whole genome analyses reveal weak signatures of population structure and environmentally associated local adaptation in an important North American pollinator, the bumble bee *Bombus vosnesenskii*. *Mol Ecol* **32**: 5479–5497.

Herbertsson L, Khalaf R, Johnson K, Bygebjerg R, Blomqvist S, Persson AS (2021). Long-term data shows increasing dominance of *Bombus terrestris* with climate warming. *Basic Appl Ecol* **53**: 116–123.

Herrera CM, Alonso C, Valverde J, Núñez A, Wood TJ (2025). The genus *Andrena* Fabricius, 1775 (Hymenoptera, Andrenidae) in a Mediterranean biodiversity hotspot: community-wide relationships with plants and description of three new species. *J Hymenopt Res* **98**: 1039–1066.

Hesselbarth MHK, Sciaini M, With KA, Wiegand K, Nowosad J (2019). landscapemetrics: an open-source R tool to calculate landscape metrics. *Ecography* **42**: 1648–1657.

Hijmans RJ (2022). Terra: Spatial Data Analysis.

Ho TAT, Pedersen JB, Aagaard A, Schou MF, Bechsgaard J, Corcoran D, *et al.* (2025). Gene Flow Disruption and Population Declines in a Soil Arthropod in Fragmented Habitats. *Mol Ecol* **34**: e17820.

Jackson JM, Pimsler ML, Oyen KJ, Koch-Uhuad JB, Herndon JD, Strange JP, *et al.* (2018). Distance, elevation and environment as drivers of diversity and divergence in bumble bees across latitude and altitude. *Mol Ecol* **27**: 2926–2942.

Jackson JM, Pimsler ML, Oyen KJ, Strange JP, Dillon ME, Lozier JD (2020). Local adaptation across a complex bioclimatic landscape in two montane bumble bee species. *Mol Ecol* **29**: 920–939.

Jaffé R, Veiga JC, Pope NS, Lanes ECM, Carvalho CS, Alves R, *et al.* (2019). Landscape genomics to the rescue of a tropical bee threatened by habitat loss and climate change. *Evol Appl* **12**: 1164–1177.

Janousek WM, Douglas MR, Cannings S, Clément MA, Delphia CM, Everett JG, *et al.* (2023). Recent and future declines of a historically widespread pollinator linked to climate, land cover, and pesticides. *Proc Natl Acad Sci* **120**: e2211223120–e2211223120.

Jha S (2015). Contemporary human-altered landscapes and oceanic barriers reduce bumble bee gene flow. *Mol Ecol* **24**: 993–1006.

Kardum Hjort C, Dudaniec RY, Olsson P, Ekroos J, Smith HG (2025). Landscape and Climate-Associated Selection in the Native and Widespread Bumblebee, *Bombus terrestris*. *Mol Ecol* **34**: e70141.

Kardum Hjort C, Paris JR, Smith HG, Dudaniec RY (2024). Selection despite low genetic diversity and high gene flow in a rapid island invasion of the bumblebee, *Bombus terrestris*. *Mol Ecol* **33**: e17212.

Karra K, Kontgis C, Statman-Weil Z, Mazzariello JC, Mathis M, Brumby SP (2021). Global land use / land cover with Sentinel 2 and deep learning. *2021 IEEE Int Geosci Remote Sens Symp IGARSS*: 4704–4707.

Katoh K, Standley DM (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol* **30**: 772–780.

Kawecki TJ, Ebert D (2004). Conceptual issues in local adaptation. *Ecol Lett* **7**: 1225–1241.

Kelemen EP, Rehan SM (2021). Conservation insights from wild bee genetic studies: Geographic differences, susceptibility to inbreeding, and signs of local adaptation. *Evol Appl* **14**: 1485–1496.

Lex A, Gehlenborg N, Strobel H, Vuilleumot R, Pfister H (2014). UpSet: Visualization of Intersecting Sets. *IEEE Trans Vis Comput Graph* **20**: 1983–1992.

- Li H (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**: 2987–2993.
- Li H (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, *et al.* (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.
- Lotterhos KE, Whitlock MC (2014). Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Mol Ecol* **23**: 2178–2192.
- Malinsky M, Trucchi E, Lawson DJ, Falush D (2018). RADpainter and fineRADstructure: Population Inference from RADseq Data. *Mol Biol Evol* **35**: 1284–1290.
- Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W-M (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**: 2867–2873.
- Marshall L, Leclercq N, Carvalheiro LG, Dathe HH, Jacobi B, Kuhlmann M, *et al.* (2024). Understanding and addressing shortfalls in European wild bee data. *Biol Conserv* **290**: 110455–110455.
- Martinet B, Rasmont P, Cederberg B, Evrard D, Ødegaard F, Paukkunen J, *et al.* (2015). Forward to the north: two Euro-Mediterranean bumblebee species now cross the Arctic Circle. *Ann Société Entomol Fr NS* **51**: 303–309.

McNeil DJ, McCormick E, Heimann AC, Kammerer M, Douglas MR, Goslee SC, *et al.* (2020). Bumble bees in landscapes with abundant floral resources have lower pathogen loads. *Sci Rep* **10**: 22306–22306.

Outhwaite CL, McCann P, Newbold T (2022). Agriculture and climate change are reshaping insect biodiversity worldwide. *Nature* **605**: 97–102.

Owen EL, Bale JS, Hayward SAL (2013). Can Winter-Active Bumblebees Survive the Cold? Assessing the Cold Tolerance of *Bombus terrestris audax* and the Effects of Pollen Feeding. *PLOS ONE* **8**: e80061.

Privé F, Luu K, Vilhjálmsson BJ, Blum MGB (2020). Performing Highly Efficient Genome Scans for Local Adaptation with R Package pcadapt Version 4. *Mol Biol Evol* **37**: 2153–2154.

R Core Team (2021). R: A Language and Environment for Statistical Computing.

Raven PH, Wagner DL (2021). Agricultural intensification and climate change are rapidly decreasing insect biodiversity. *Proc Natl Acad Sci* **118**: 1–6.

Rouault E, Warmerdam F, Schwehr K, Kiselev A, Butler H, Łoskot M, *et al.* (2025). GDAL.

Rousset F (1997). Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* **145**: 1219–1228.

Shapiro SS, Wilk MB (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika* **52**: 591–611.

Shumate A, Salzberg SL (2021). Liftoff: accurate mapping of gene annotations. *Bioinformatics* **37**: 1639–1643.

Silva SE, Seabra SG, Carvalheiro LG, Nunes VL, Marabuto E, Mendes R, *et al.* (2020). Population genomics of *Bombus terrestris* reveals high but unstructured genetic diversity in a potential glacial refugium. *Biol J Linn Soc* **129**: 259–272.

Sirami C, Gross N, Baillo AB, Bertrand C, Carrié R, Hass A, *et al.* (2019). Increasing crop heterogeneity enhances multitrophic diversity across agricultural regions. *Proc Natl Acad Sci* **116**: 16442–16447.

Soroye P, Newbold T, Kerr J (2020). Climate change contributes to widespread declines among bumble bees across continents. *Science* **367**: 685–688.

Sousa P de, Henriques A, Silva SE, Carvalheiro LG, Smagghe G, Michez D, *et al.* (2023). Genomic Patterns of Iberian Wild Bees Reveal Levels of Diversity, Differentiation and Population Structure, Supporting the “Refugia within Refugia” Hypothesis. *Divers 2023 Vol 15 Page 746* **15**: 746–746

Sousa P, Theodorou P, Martins AC, Aguiar AJC, Shibata A, Kahnt B (2025). Integrating Population Genomics and Historical Occurrence Data to Assess the Impact of Land-Use Changes on Centris Oil Bees in the Brazilian Cerrado. *Mol Ecol* **34**: e70153.

Szczepko-Morawiec K, Wiśniowski B, Motyka E, Celary W, Kruk A (2024). Ecological amplitude and indication potential of mining bees (*Andrena* spp.): a case study from the post-agricultural area of the Kampinos National Park (Poland). *Sci Rep* **14**: 9738.

Theodorou P, Baltz LM, Paxton RJ, Soro A (2021). Urbanization is associated with shifts in bumblebee body size, with cascading effects on pollination. *Evol Appl* **14**: 53–68.

Theodorou P, Radzevičiūtė R, Kahnt B, Soro A, Grosse I, Paxton RJ (2018). Genome-wide single nucleotide polymorphism scan suggests adaptation to urbanization in an important pollinator, the red-tailed bumblebee (*Bombus lapidarius* L.). *Proc R Soc B Biol Sci* **285**: 20172806.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, *et al.* (2012). Primer3—new capabilities and interfaces. *Nucleic Acids Res* **40**: e115–e115.

Viana CM, Freire D, Abrantes P, Rocha J (2021). Evolution of Agricultural Production in Portugal during 1850–2018: A Geographical and Historical Perspective. *Land* **10**: 776.

Weissel N, Mitesser O, Liebig J, Poethke H-J, Strohm E (2006). The influence of soil temperature on the nesting cycle of the halictid bee *Lasioglossum malachurum*. *Insectes Sociaux* **53**: 390–398.

Wood TJ, Müller A, Praz C, Michez D (2023). Elevated rates of dietary generalization in eusocial lineages of the secondarily herbivorous bees. *BMC Ecol Evol* **23**: 67.

Wyman LM, Richards MH (2003). Colony social organization of *Lasioglossum malachurum* Kirby (Hymenoptera, Halictidae) in southern Greece. *Insectes Sociaux* **50**: 201–211.

Yeaman S (2022). Evolution of polygenic traits under global vs local adaptation. *Genetics* **220**.

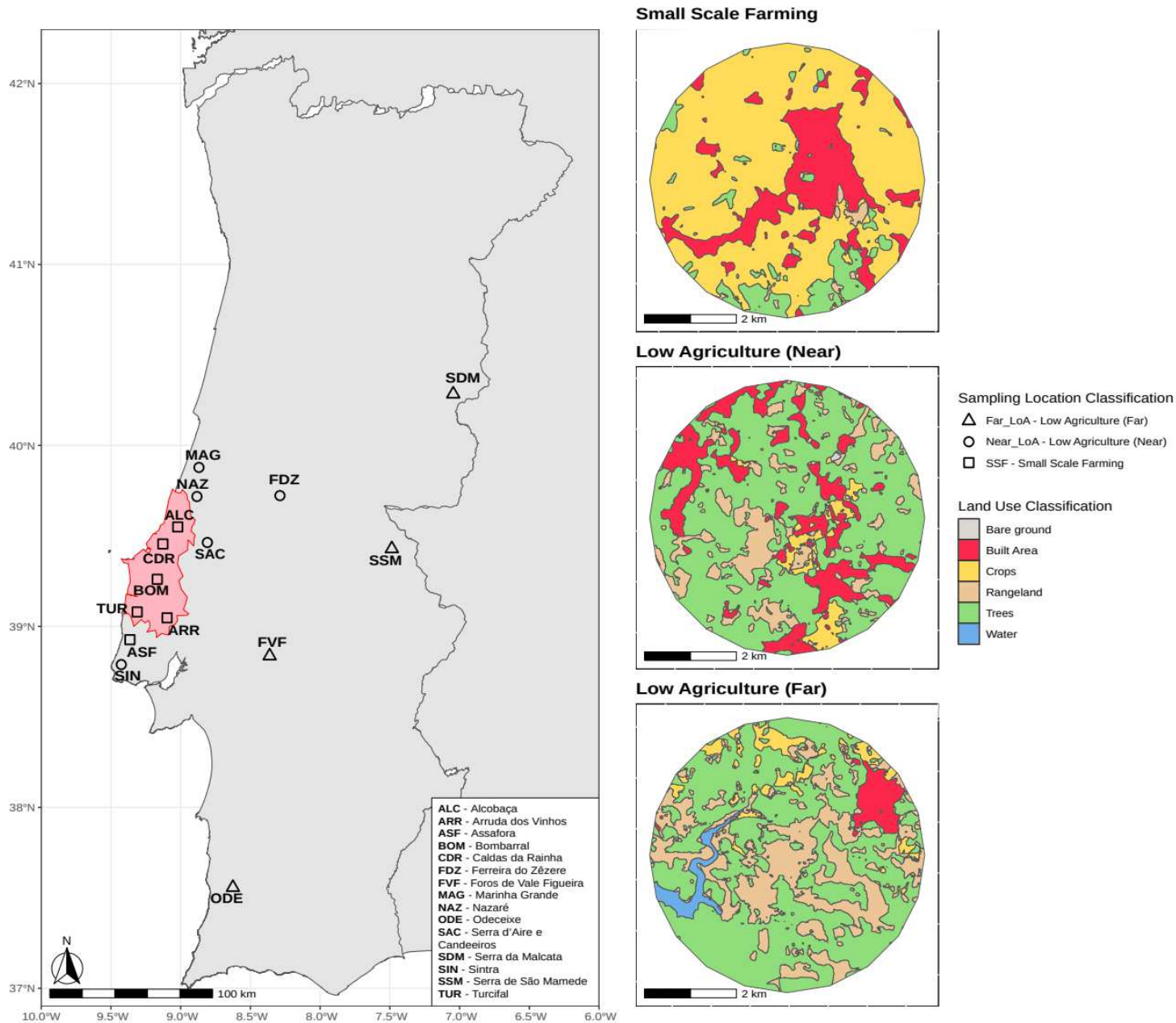
Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**: 3326–3328.

682

683

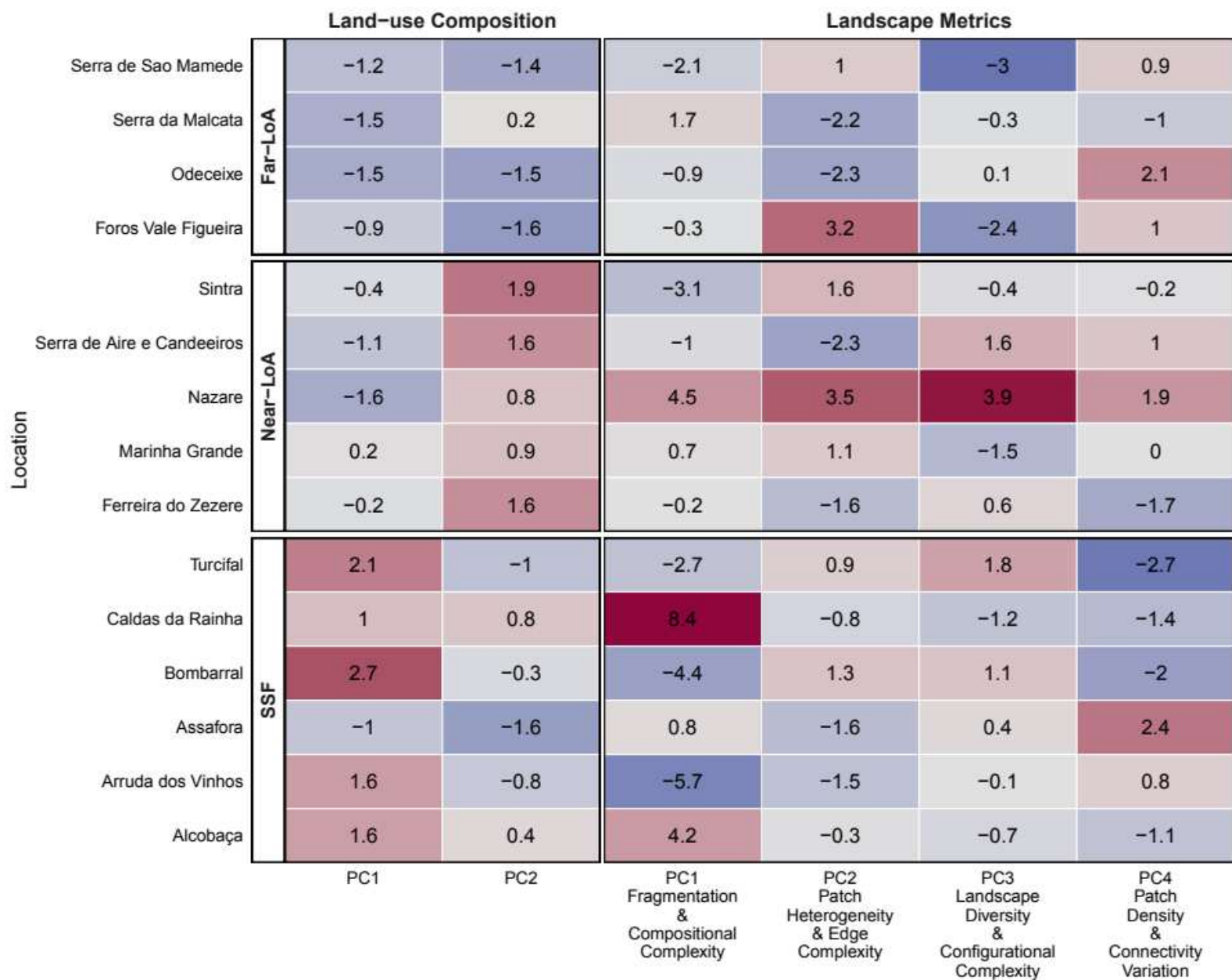
684

685 Figures legends



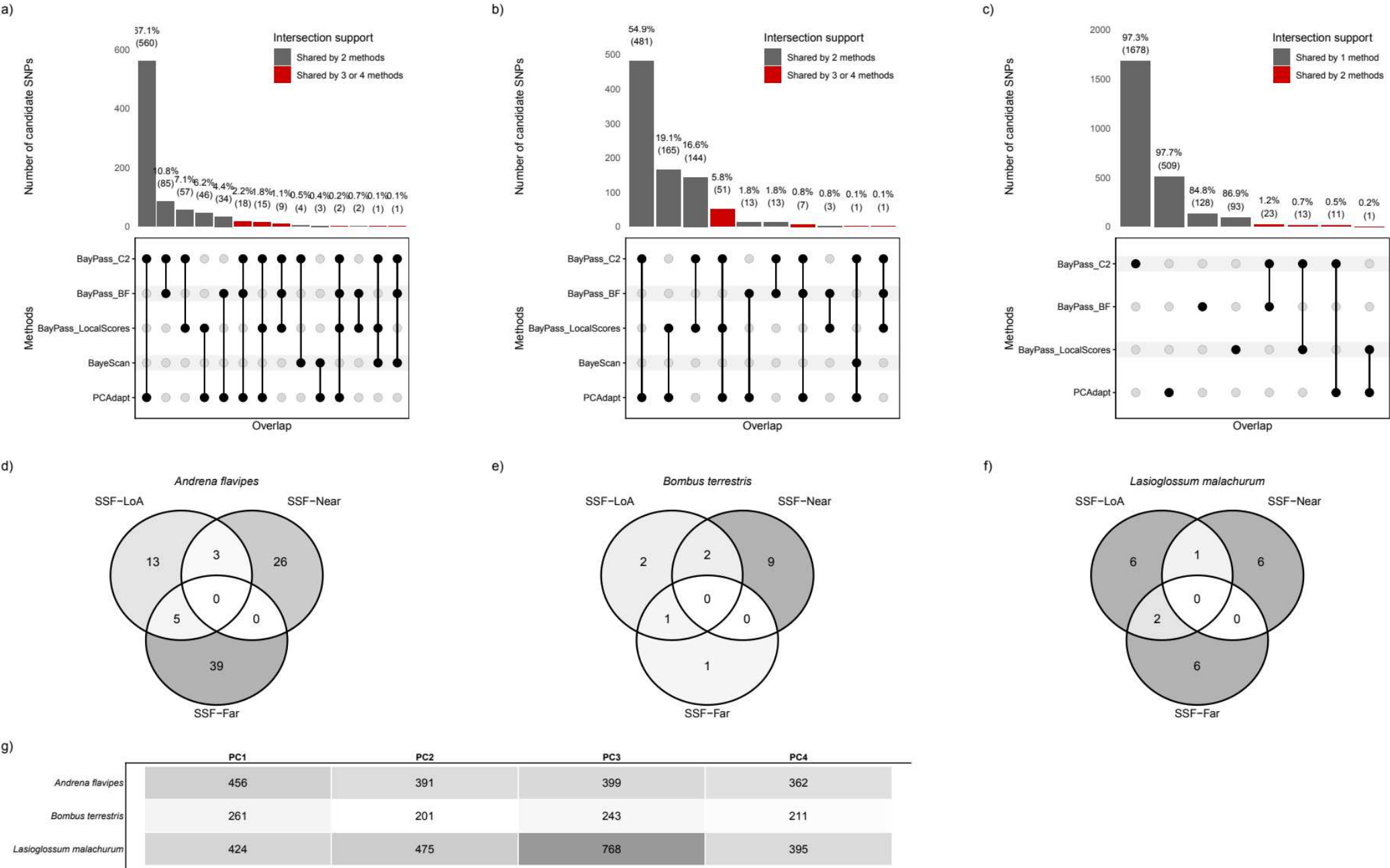
686 **Figure 1.** Sampling locations and landscape-category classification. Left panel:
 687 geographic distribution of the 15 sampling locations across Portugal. Symbols indicate
 688 landscape category: squares, small-scale farming (SSF) landscapes; circles, near low-

689 agricultural (Near_LoA) landscapes; triangles, far low-agricultural (Far_LoA)
690 landscapes. Right panel: representative land-use composition maps for each
691 landscape category (SSF, Near_LoA, and Far_LoA), illustrating the dominant land-use
692 classes used for landscape classification. Colors indicate land-use classes (bare
693 ground, built area, crops, rangeland, trees, and water), and scale bars indicate spatial
694 extent.



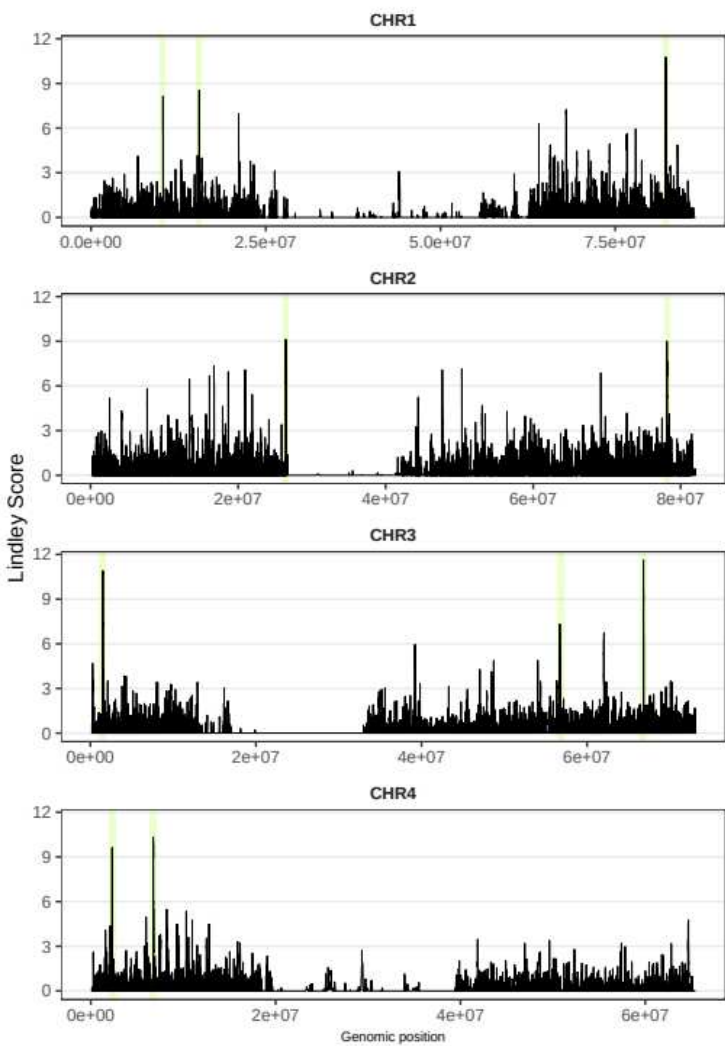
695
696 **Figure 2.** PCA scores of land-use composition and landscape metrics axes across
697 sampling locations. Heat map showing standardized PCA scores for each sampling
698 location classified as Far_LoA, Near_LoA, or SSF. The left panel displays scores from
699 the two principal components derived from land-use composition, whereas the right

700 panel displays scores for the first four principal components derived from landscape
701 metrics. Positive values are shown in red and negative values in blue.

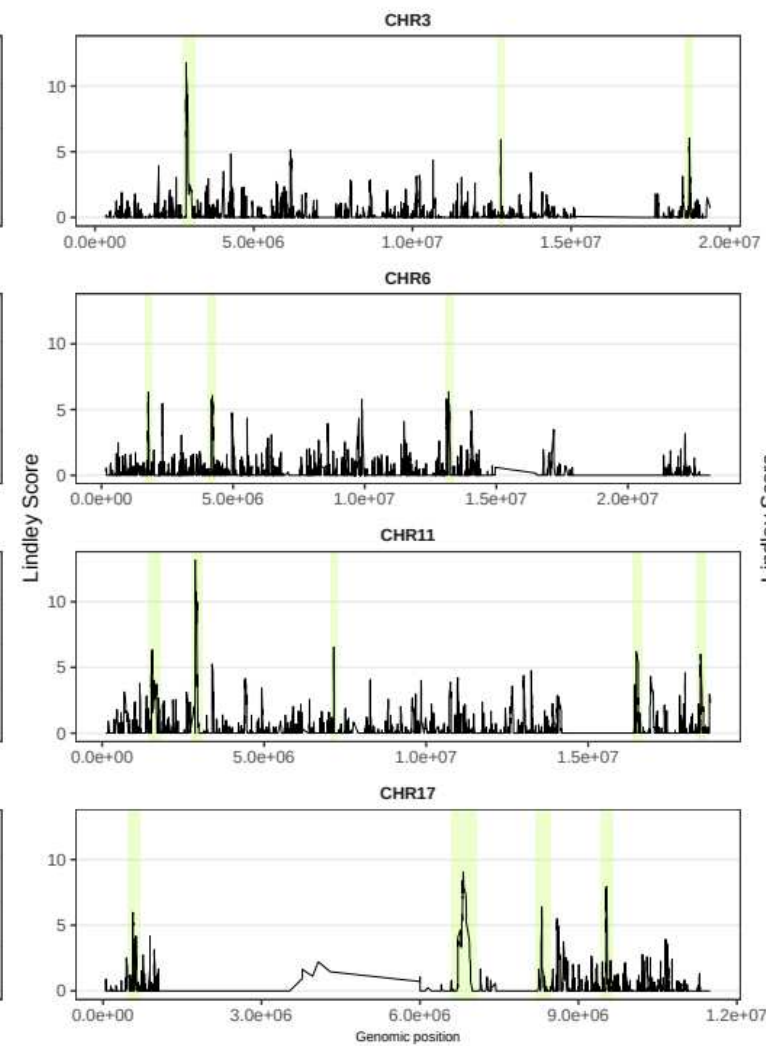


703 **Figure 3.** Summary of selection detection and GEA analyses across the three studied species. (a-c) Overlap
704 among candidate SNPs identified by different selection detection methods. a) *Andrena flavipes*, b) *Bombus*
705 *terrestris*, and c) *Lasioglossum malachurum*. Vertical bars show the number of SNPs, with percentages above
706 bars. Connected filled circles indicate the methods contributing to each intersection. The red bars represent the
707 outliers used as conservative candidates. (d-f) Venn diagrams showing the overlap of conservative candidate
708 SNPs among the three land-use contrasts (SSF x LoA, SSF x Near_LoA, and SSF x Far_LoA). d) *Andrena flavipes*, e)
709 *Bombus terrestris*, and f) *Lasioglossum malachurum*. g) Distribution of decisive GEA associations across the four
710 landscape metrics PCA axes: PC1 (fragmentation and compositional complexity), PC2 (patch heterogeneity and
711 edge complexity), PC3 (landscape diversity and configurational complexity), and PC4 (patch density and
712 connectivity variation).

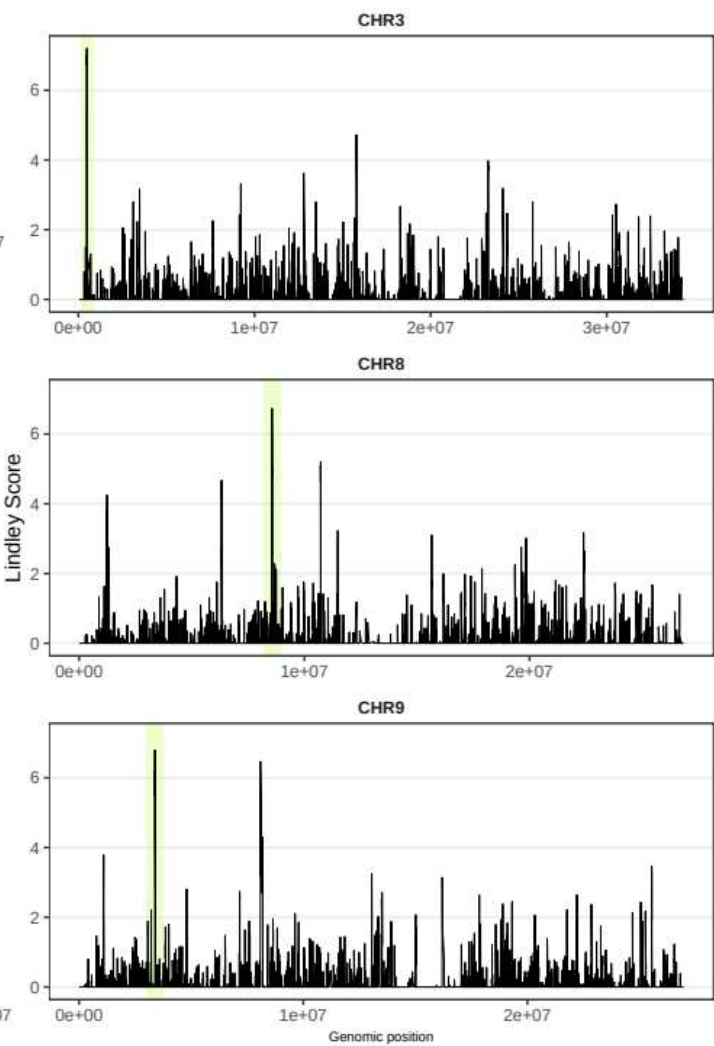
a) *Andrena flavipes* (SSF x Far_LoA)



b) *Bombus terrestris* (SSF x Near_LoA)



c) *Lasioglossum malachurum* (SSF x LoA)



715 **Figure 4.** Genomic distribution of significant local score windows in the contrasts showing the strongest evidence
716 of differentiation for each species. a) *Andrena flavipes* (SSF x Far_LoA), b) *Bombus terrestris* (SSF x Near_LoA), and
717 c) *Lasioglossum malachurum* (SSF x LoA). Black bars represent Lindley scores calculated from BayPass C2 p-
718 values along each represented chromosome. Green shaded regions indicate genomic windows identified as
719 significant by the BayPass local score approach.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [AppendixI.docx](#)
- [SupplemenrtaryTableslegends.pdf](#)
- [TableS16.pdf](#)
- [TableS19.pdf](#)
- [TableS17.pdf](#)
- [TableS15.pdf](#)
- [TableS20.pdf](#)
- [TableS11S14.pdf](#)
- [TableS18.pdf](#)
- [TableS1S10.pdf](#)
- [SupplementaryFigures.tar.gz](#)
- [TableS21S32.pdf](#)